final report

Project code: P.PIP.0751

Prepared by: John Langbridge
Teys Australia Pty Ltd

Date published: 16 May 2018

PUBLISHED BY
Meat and Livestock Australia Limited
Locked Bag 1961
NORTH SYDNEY NSW 2059

Training the E+V grading camera against Australian meat grading standards

This is an MLA Donor Company funded project.

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government and contributions from the Australian Meat Processor Corporation to support the research and development detailed in this publication.

Abstract

Both the recent ACCC inquiry and the Senate Rural Affairs have highlighted the need to develop more objective methods for assessing carcases value. In Australia, meat quality attributes are inspected by professionally trained assessors. Scientific studies have shown that manual grading may lead to inconsistencies and variations.

The use of a calibrated grading camera, as used in the United States, is at least as accurate when measuring each quality trait of the exposed rib-eye muscle as the plant grader when compared to the Expert (Meat Standards Australia) Grader. Overall the correlations were good and the camera prediction model appears to perform at least as accurately when compared to the expert grader as the current plant graders. It also produced less variable results. Essentially it showed that it had promise. Furthermore a permanent record of the grading outcome is available whenever there are disputes.

There is now sufficient information to support a recommendation to conduct of a full trial with the aim of seeking formal AUSMEAT approval as an aid to carcase grading under Australian conditions. This trial will additionally address inter camera variability and repeatability of camera outputs.

Executive summary

Both the recent ACCC inquiry and the Senate Rural Affairs have highlighted the need to develop more objective methods for assessing a carcases value. In Australia, meat quality attributes are inspected by professionally trained assessors. Scientific studies have shown that manual grading may lead to inconsistencies and variations.

The project shows that the grading camera (E+V) is at least as accurate in measuring each trait as the plant grader when compared to the Expert (MSA) Grader. It also produced less variable results than the commercial graders.

It was found to be more accurate in measuring marbling score, the camera accounted for 68% of the variation in Expert Grader score. Plant graders were only 61%. For fat thickness there was a big variation between expert & plant graders (only 41% correlation). The camera was more closely correlated with expert grader (69%). Note that the camera measured the average thickness rather than at the defined point as per the MSA requirements. For the rib eye area there was not as much difference. The camera was 78% correlated with Expert grader; with plant graders at 76%.

A regression model was developed for meat colour, fat colour & marbling. This was then used to help “train” the camera. For meat colour the camera classified meat colour 61% correctly. Plant graders were only at 36% agreement with expert graders. Plant graders tended to assign a more favourable lean colour than the expert. For fat colour the camera classified 71% of samples correctly when compared to expert grader; whereas plant graders were 43% in agreement. There was a large difference in the distribution of the plant grader scores vs expert grader scores.

For the MSA marbling score the camera correctly classified 67%. Similar results for plant graders (65%). However plant graders tended to award higher marbling scores on average than the expert grader.

Overall the correlations were good and the camera prediction model appears to perform at least as accurately when compared to the expert grader than the current plant graders. Any variations in the individual attribute measurements did not affect the overall grading outcome.

The next step in the process will be to seek AUSMEAT approval. That will require that the potential issue of inter camera variability will need to be assessed and a system of calibration agreed.

A new project is envisaged to generate the necessary information required for AUSMEAT approval, with both expert and commercial graders grading the same carcases independently and any results with large disagreements within each group of graders discarded to produce a final data set that will be used to support the application for approval. Descriptive statistics for the carcasses to be finalised in a trial protocol, i.e.; the mean, standard dev, max, min of carcass weight, rib fat, EMA, MSA marble etc. and submitted to both MLA and AUSMEAT prior to the new trial commencing.

Table of contents

1 Background

Both the recent ACCC inquiry and the Senate Rural Affairs have highlighted the need to develop more objective methods for assessing a carcases value. The current Rural R&D for Profit project covering Advanced Measurement Technologies for globally competitive Australian meat (ALM Tech) is also looking to develop these objective methods. One of the more common areas of dispute with producers is the grading of beef carcases.

Traditionally, meat quality attributes are inspected by professionally trained assessors. Even using professionally trained graders, visual inspection of meat quality may lead to inconsistencies and variations (Cheng et al., 2015). Furthermore, manual inspection is labour intensive, time-consuming, tedious and costly, and can be influenced by physiological factors resulting in subjective and inconsistent evaluation (Valous et al., 2016).

Machine grading is conducted in the US by the USDA Agricultural Marketing service. There are two technologies that are currently approved after validation against the USDA grading standards. They are CVS (Computer Vision System; RMS Research Management Systems, USA, Inc., Fort Collins, CO) and VBG2000 (E+V Technology, Oranienburg, Germany). The instruments incorporate several variables including the amount, size, and distribution of fat (marbling) present within the exposed ribeye, as well as variables of lean and fat colour (Mafi et al 2014).

2 Project objectives

2.1 Train the E+V Camera

Train the E + V Camera against the Australian grading standards for Meat Colour, Fat Colour, Marbling, Eye Muscle Area, and Rib Fat Measurement,

2.2 Identify the approval path

Seek AUSMEAT/MSA agreement of a process to approve for use under Australian grading conditions the E + V camera in AUSMEAT/MSA accredited beef processing plants.

3 Methodology

The project seeks to train (develop the algorithm) for the E + V cold grading camera currently used by the USDA, to assess Australian meat grading standards as specified by AUSMEAT and Meat Standards.

This will be achieved by assessing the same carcases as the normal MSA qualified company graders, the MSA expert graders and -one E + V camera. This was be conducted at one abattoir initially which will cover most grading outcomes. The abattoir was the Teys Australia plant at Wagga Wagga.

Data was collected during routine grading on three consecutive days

All sides that were presented and ribbed well enough for evaluation were imaged with the VBG2000 grading camera, the same one that is used in the US.

On the second day, it is estimated that 98.8% (1,867 out of 1,890) of presented sides were successfully imaged.

MSA expert graders collected data on the first two days. The plant grader was data assembled for all carcasses. A small number of cattle do not have MSA traits because they were not MSA eligible

A team of students from Texas Tech University trained in US grading, evaluated USA grading traits. On the first day (Monday) one team member evaluated quality traits and the other team member evaluated yield grade traits. On the second and third day (Tuesday and Wednesday), both team members evaluated quality. This allowed for averaging of the subjective data which increased the precision of those evaluations, which made them more strongly related to the camera data.

4 Results

4.1 Data Set

The carcasses graded for this project included extreme diversity for:

- Carcass weight
- Fatness
- Muscularity
- Marbling
- Lean color
- Fat color
- Genetics
  - Primarily *Bos Taurus*
  - Some high percentage *Bos indicus*
- Age
- Diet
  - Grass-fed
  - Grain-fed
- 1,893 head with image data for both sides
  - 712 with expert MSA marbling scores
  - 1,773 with on-line MSA marbling scores
  - 1,887 with TTU expert USA marbling scores
    - 1,691 with panel (mean of two experts) marbling scores
    - 196 with individual expert marbling scores
- 730 head with image data for the left side only
  - 451 with expert MSA marbling scores
  - 668 with on-line MSA marbling scores
  - 729 with TTU expert USA marbling scores
    - 697 with panel (mean of two experts) marbling scores
    - 32 with individual expert marbling scores
- 13 head with image data for the right side only

4.2 Factors that affect the quality and usefulness of data from the VBG2000.

4.2.1 Carcass side identification

The machine grading was conducted independently of the normal grading process which captures the carcase bar code.

4.2.2 Ribbing

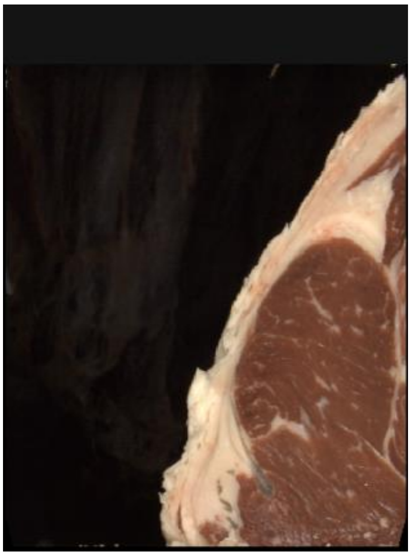
For the camera to properly evaluate the traits of interest the camera nose must be placed on a flat surface. This is not possible if the carcass is mis-ribbed. The cut must be made with a slight upward angle so that the cut is smooth and straight for a distance of at least 24 cm to allow the camera nose to be properly positioned. Failure to do so results in:

- The camera nose being too far away from the object of interest
- Underestimation of ribeye area
- Underestimation of the brightness of lean colour (false classification as a dark cutter)

Example carcass (170760) with both sides properly ribbed

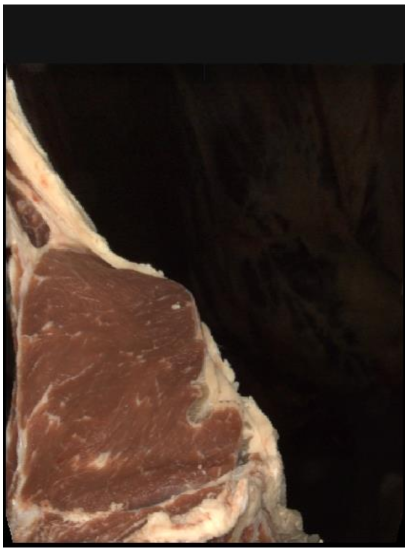
Side 1 (left side of carcass)

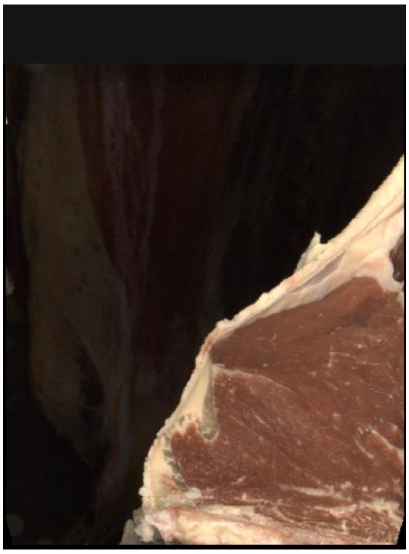
Side 2 (right side of carcass)



Example carcass (160913) with both sides scooped because the ribbing cut was made perpendicular to the backbone and resulted in the cut intersecting the 12th rib. Therefore, the camera was too far from the object of interest.

Side 1 (left side of carcass) Side 2 (right side of carcass)



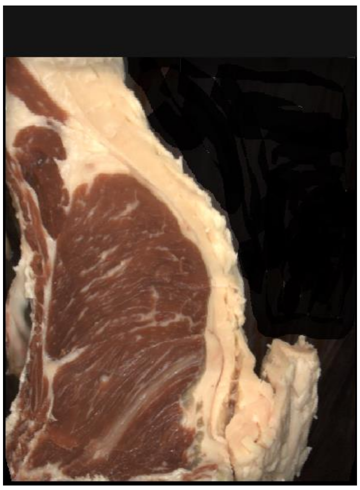


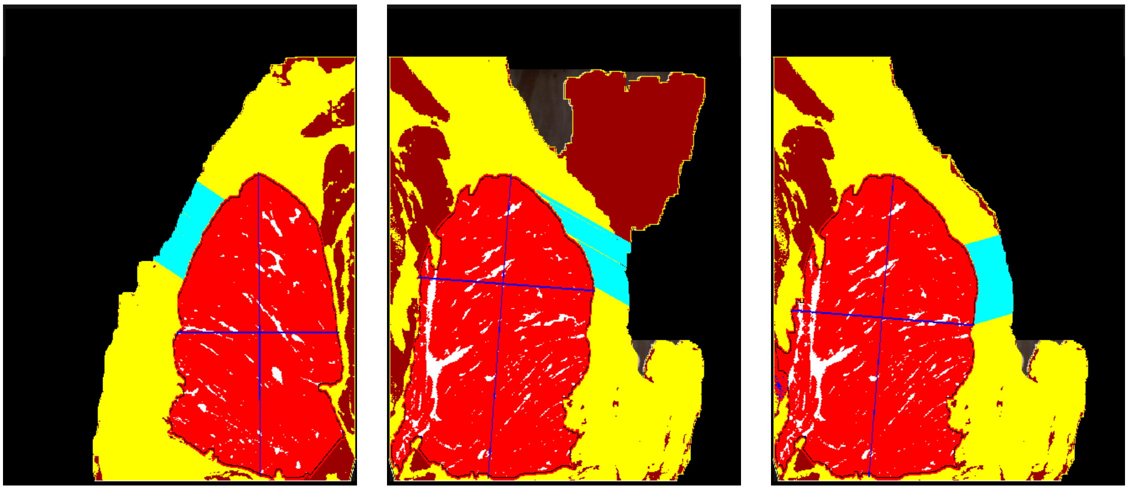
4.2.3 Carcass side orientation and spacing.

A side cannot be imaged if it is facing away from the camera operator. Moreover, if a side is facing the wrong direction, it is virtually impossible to keep that side from interfering with the operators ability to image adjacent sides.

This is an example of what happens when there is not enough space between adjacent sides. For 911's left side, the image analysis could not distinguish the object of interest from the plate of the preceding side (910's right side).

911-R 911-L 911L fixed





4.2.4 Averaging out errors.

Evaluating both sides of the carcass and taking the average of the two sides for each prediction will provide the industry with the most useful data. In the USA, the grading system is based on taking the marbling score from the “high” side (i.e., the grader determines the marbling score for whichever side of the carcass that the grader deemed to have the highest marbling score). The same approach is taken with instrument grading. In the USA, the percentage of carcasses grading U.S. Choice or higher would be ~ 7% lower if we only imaged one side of the carcass. If a plant only imaged one side, that would be equivalent to an average economic loss of $6 for each head processed. In Australia the current MSA practice is to measure traits on only side only. The other side is used only if required due to damage.

4.3 Results

4.3.1 Evaluation of traits with USA-based prediction models

One-half of the carcasses were assigned to the calibration data set which was used to develop the model and one-half of the carcasses were assigned to the prediction data set which was used to evaluate the model. Carcasses were ranked by TTU expert panel for marbling score and alternately assigned to the calibration and prediction data sets. Other attributes were also measured and recorded against each carcase.

4.3.1.1 Marbling score

Both the VBG2000 marbling score and TTU expert marbling score accounted for 68% of the variation in expert MSA marbling score. In comparison, the plant graders’ MSA marbling score accounted for 60.6% of the variation in expert MSA marbling score. VBG2000 marbling score was more strongly related (R^2 0.820 vs 0.701 and 0.682) to TTU expert marbling score than it was to either the plant graders’ marbling score or the expert MSA marbling score. This was likely due to the averaging of two TTU expert evaluations for the TTU panel on the second and third day of data collection.

Table 1. Coefficient of determination (R^2) for relationship between various measures of marbling score.

	VBG2000 marbling score	TTU expert USA marbling score	Plant Graders' MSA marbling score	Expert MSA marbling score
VBG2000 marbling score		0.820	0.701	0.682
TTU expert USA marbling score	0.820		0.725	0.681
Plant Graders' MSA marbling score	0.701	0.725		0.606
Expert MSA marbling score	0.682	0.681	0.606	

4.3.1.2 Fat thickness

The VBG2000 predicts preliminary yield grade (PYG), which can be mathematically converted to fat thickness, and adjusted preliminary yield grade (ADJ), which can be mathematically converted to adjusted fat thickness. VBG2000 adjusted fat thickness was much more strongly correlated to MSA expert rib fat depth than was the plant graders' fat depth (R^2 = 0.689 vs 0.411).

Table 2. Coefficient of determination (R^2) for relationship between various measures of fat thickness.

	VBG2000 adjusted fat thickness	TTU expert adjusted fat thickness	Plant Graders' fat depth	Expert fat depth
VBG2000 adjusted fat thickness		0.727	0.414	0.689
TTU expert adjusted fat thickness	0.727		0.469	0.739
Plant Graders' fat depth	0.414	0.469		0.411
Expert fat depth	0.689	0.739	0.411	

4.3.1.3 Ribeye area.

VBG2000 ribeye area was slightly more strongly correlated to MSA expert eye muscle area than was the plant graders' eye muscle area measurement (R^2 = 0.782 vs 0.756).

Table 3. Coefficient of determination (R^2) for relationship between various measures of longissimus muscle area.

	VBG2000 ribeye area	TTU expert ribeye area	Plant Graders' eye muscle area	Expert eye muscle area
VBG2000 ribeye area		0.742	0.683	0.782
TTU expert ribeye area	0.742		0.711	0.760
Plant Graders' eye muscle area	0.683	0.711		0.756
Expert eye muscle area	0.782	0.760	0.756	

4.3.2 Evaluation of traits requiring development and evaluation of novel prediction models. (Accuracy of prediction of MSA expert with VBG2000)

For all of the following traits, the same approach was taken. One-half of the carcasses were assigned to the calibration data set which was used to develop the model and one-half of the carcasses were assigned to the prediction data set which was used to evaluate the model.

4.3.2.1 Meat Colour

Because MSA Meat Colour score is a different scale than North American systems, a prediction model had to be developed and evaluated. The regression model was trained to predict the MSA expert Meat Colour score. Because MSA expert Meat Colour score contained both quantitative and qualitative scores, it was necessary to convert the scores to a quantitative scale for regression. Preliminary analysis showed that VBG ribeye colour was linear across the scores of 2 through 7 and that 1B and 1C were linearly equivalent to approximately 1 and 1.5, respectively. Therefore, all MSA expert Meat Colour scores of 1B and 1C were converted to 1 and 1.5, respectively. The same approach was taken for the on-line grader's data.

The regression model was trained to predict the MSA expert scores for the various Australian Meat Quality attributes

The prediction equation was developed with the calibration data set and the top panel of table 4 and shows the evaluation of the prediction equation with the prediction data set. Most (60.7%) samples were correctly classified and only 2.2% were misclassified by more than a 1 class error. The bottom panel shows the level of agreement between the plant grader meat colour score and MSA expert meat colour score for the same set of 364 carcasses. The level of agreement was 36.0% and 10.7% were misclassified by more than a 1 class error. Plant graders tended to assign carcasses a more favourable lean colour score than did the expert.

Table 4. Accuracy of prediction of MSA expert meat colour score with VBG meat color score.

cal_pred = 2_Pred
Sides = BothSides

221	60.7	% agreement
135	37.1	% 1 class error
8	2.2	% 2 class error
	0.0	% 3 class error

VBG MeatColour score	MSA Expert MeatColour score 1B	1C	2	3	4	5	6	7	Grand Total
1B		1	2						3
1C	4	15	12	1					32
2		9	44	35					88
3		5	25	86	23				139
4				20	70	3			93
5					2	6	1		9
6									0
7									0
Grand Total	4	30	83	142	95	9	1	0	364

cal_pred = 2_Pred
Sides = BothSides

131	36.0	% agreement
194	53.3	% 1 class error
38	10.4	% 2 class error
1	0.3	% 3 class error

Plant Graders meat_colour score	MSA Expert MeatColour score 1B	1C	2	3	4	5	6	7	Grand Total
1B									0
1C	3	13	17	8					41
2	1	16	55	76	25				173
3		1	11	58	69	3	1		143
4					1	2			3
5						4			4
6									0
7									0
Grand Total	4	30	83	142	95	9	1	0	364

4.3.2.2 Fat Colour

The prediction equation was developed with the calibration data set and the top panel of table 5 shows the evaluation of the prediction equation with the prediction data set. Most (70.9%) samples were correctly classified and only 2.7% were misclassified by more than a 1 class error. The bottom panel shows the level of agreement between the plant grader meat colour score and MSA expert meat colour score for the same set of 364 carcasses. The level of agreement was 43.4% and 5.8% were misclassified by more than a 1 class error. The distribution of fat colour scores differed between plant graders and the expert. The expert's scores were bimodal and included only 7% with fat colour scores of 1. In contrast, 38% of the plant grader's scores were 1.

Table 5. Accuracy of prediction of MSA expert fat colour score with VBG fat colour score.

cal_pred = 2_Pred
Sides = BothSides

258	70.9	% agreement
96	26.4	% 1 class error
9	2.5	% 2 class error
1	0.3	% 3 class error

VBG FatColour score	MSA Expert FatColour score 0	1	2	3	4	5	6	7	Grand Total
0	193	17	1						211
1	8	6	6	1					21
2	2	2	15	23					42
3	1	1	15	38	8				63
4				7	1	5			13
5				4	2	3	2		11
6						1	2		3
7									0
Grand Total	204	26	37	73	11	9	4	0	364

cal_pred = 2_Pred
Sides = BothSides

158	43.4	% agreement
186	51.1	% 1 class error
19	5.2	% 2 class error
1	0.3	% 3 class error

Plant Graders Fat_colour score	MSA Expert FatColour score 0	1	2	3	4	5	6	7	Grand Total
0	93	6	2						101
1	108	13	10	9					140
2	3	7	11	25	2	1			49
3			14	33	8	1			56
4				4					4
5				2	1	7	1		11
6							1		1
7							2		2
Grand Total	204	26	37	73	11	9	4	0	364

4.3.2.3 Ausmeat marbling

The regression model was trained to convert VBG marbling score to MSA expert AUS marbling score. When the model was applied to the prediction data set, most (67.3%) samples were correctly classified and only 0.5% were misclassified by more than a 1 class error. This was similar to the level of agreement between plant graders and the MSA expert. But, plant graders tended to award higher marbling scores than the expert.

Table 5. Accuracy of prediction of MSA expert Ausmeat Marbling score with VBG marbling score.

cal_pred = 2_Pred
Sides = BothSides

245	67.3	% agreement
117	32.1	% 1 class error
2	0.5	% 2 class error
0	0.0	% 3 class error

VBG AUS marbling score	MSA Expert AUS marbling score 0	1	2	3	4	5	Grand Total
0	13	13					26
1	20	177	42	1			240
2		22	50	6			78
3			8	3	3		14
4			1	2	1		4
5					1	1	2
Grand Total	33	212	101	12	5	1	364

cal_pred = 2_Pred
Sides = BothSides

238	65.4	% agreement
121	33.2	% 1 class error
5	1.4	% 2 class error
0	0.0	% 3 class error

Plant graders' AUS marbling score	MSA Expert AUS marbling score 0	1	2	3	4	5	Grand Total
0	23	24					47
1	10	136	17				163
2		51	71	5	1		128
3		1	11	5	1		18
4			2	1	2		5
5				1	1	1	3
Grand Total	33	212	101	12	5	1	364

4.3.3 Common traits

4.3.3.1 US marbling score.

The camera is already programmed to assess for US marbling score.

4.3.3.2 Rib Fat depth.

The camera is already programmed to assess for rib fat depth but the measurement is in imperial and will need to be converted. The other issue is that the fat depth is an average rather than a measurement at a specific point

4.3.3.3 Eye Muscle Area.

The camera is already programmed to assess for Eye Muscle Area but the measurement is in imperial and will need to be converted

5 Discussion/recommendations

5.1 Next Steps

5.1.1 AUSMEAT approval

The algorithms have now been developed. The issue before approval can be sought involve the analysis of inter-equipment variability and the repeatability of the camera when assessing multiple exposures to the same camera. A further trial will need to be conducted involved at least three grading cameras.

After discussion with AUSMEAT the proposed trial will be:

<u>Trial plant</u>

- Wagga Wagga

<u>Equipment</u>

- 3 E+V grading cameras installed at existing stationary grading stand (1 spare)
- As an output of this project the cameras have been programed to output, meat and fat colour. The current algorithm that outputs Marbling (MSA), Rib Eye Area, and fat depth (average rather than spot measurement) is adequate
- 3 barcode scanners to read carcase tickets

<u>Personnel</u>

- 3 skilled users of the E+V camera
- MSA expert graders (2)
- Commercial graders (existing)

<u>Method</u>

- Over three to five days, grade 1000 carcases using:
    - the three cameras;
    - MSA expert graders; and
    - commercial (Teys) graders.
- Ensure that there are animals representative of most possible grading outcomes (such as bos-indicus cross)
    - Noting that the marbling score for MSA is the US system
- Assess repeatability over 50 carcases per day (150 in total)
- Parameters being assessed are Rib eye area, Meat colour, fat colour, rib-eye fat depth (check average vs spot measurement), and marbling (MSA)
    - Australian MSA marbling score is not yet included in the cameras algorithm
- Collate and statistically analyse the data.

    - Compare to commercial and expert grader results for the parameters being measured
    - Assess significance of any correlations
    - Assess repeatability of the individual cameras
    - Assess inter-camera variability
- Calibrate the camera using E+V method
- Using “Oscap” standardised images or images selected from the trial, test alternate “Australianised” calibration system
    - Images will be provided through project.

<u>Calibration</u>

- Assess and make recommendation about the current calibration system
    - See Appendix
- Assess and make recommendations about the “Oscap” based calibration system or alternate.

<u>Data Analysis</u>

- Data will be presented in a form that has been use by the ALMTech group in other work e.g. scatter plots, data range, regression, Root Mean Square Error, and R^2,

<u>Application</u>

- Develop application to submit to AUSMEAT for approval of the E+V camera
- Send to Rod Polkinghorn to proof the application
- Send final application to AUSMEAT

5.1.2 Bloom time

Bloom time before imaging was very long and somewhat inconsistent. This lead to rib-eyes that were weeping copious amounts of purge before imaging.

To the human eye, marbling is easier to see when the ribeye has bloomed, the VBG2000 does not have that restriction and in fact is impeded if the surface to be analysed has too much sheen as would be caused by purge. The VBG2000 marbling score is very consistent from 1 to 15 minutes of bloom and then begins to drop.

6 Bibliography

Cheng, W., Cheng, J.-H., Sun, D.-W. and Pu, H. (2015), Marbling Analysis for Evaluating Meat Quality: Methods and Techniques. Comprehensive Reviews In Food Science And Food Safety, 14: 523–535.

Mafi, G., Harsh, B., Scanga, J,. Review of Instrument Augmented Assessment of USDA Beef Carcass Quality Grades. 67th Annual Reciprocal Meat Conference, American meat Science Association 2014.

Valous, N.A, Zheng, L., Sun, D.W., Tan, J (2016) Quality evaluation of meat Cuts, COMPUTER VISION TECHNOLOGY FOR FOOD QUALITY EVALUATION, Ch 7, 175-193.

7 Appendix

7.1 VBG 2000 calibration

True calibration is done by the E+V service technicians. Calibration is checked daily with the system check process. A series of cards are imaged by the system to check to make sure that the geometry and color are still within tolerance. Although in frequent, it is possible to knock the camera out of position. Likewise, an error could occur in light intensity. Most plants in North America contract with E+V to conduct quarterly service checks. When E+V technicians do their service checks, they make a new reference image.

System check cards

Small Medium Large

Example of imaging a marbling card

High Medium Low

Imaging a marbling card